

1/5/7 (Item 7 from file: 351)
DIALOG(R) File 351:Derwent WPI
(c) 2006 The Thomson Corp. All rts. reserv.

013840868 **Image available**
WPI Acc No: 2001-325081/ 200134
XRPX Acc No: N01-234468

Voice processing apparatus processes speech recognition result and
homonym demanded for incorrect input voice and outputs to machine
translator for reprocessing

Patent Assignee: SONY CORP (SONY)
Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2001100788	A	20010413	JP 99277742	A	19990930	200134 B

Priority Applications (No Type Date): JP 99277742 A 19990930

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 2001100788	A	16	G10L-015/22	

Abstract (Basic): JP 2001100788 A

NOVELTY - The speech processing apparatus connected to machine translator (2) performs speech recognition process of the input voice followed by natural language processing. A homonym which is demanded during the processing for an unrecognized input word is also provided to the machine translator for processing the input voice again.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) Voice processing method;
- (b) Recording medium

USE - For high level language processing.

ADVANTAGE - The speech recognition is again performed depending on the homonym demand, hence high level natural language processing is enabled.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of voice translator. (Drawing includes non-English language text).

Machine translator (2)
pp; 16 DwgNo 1/15

Title Terms: VOICE; PROCESS; APPARATUS; PROCESS; SPEECH; RECOGNISE; RESULT;
DEMAND; INCORRECT; INPUT; VOICE; OUTPUT; MACHINE; TRANSLATION;
REPROCESSING

Derwent Class: P86; T01; W04

International Patent Class (Main): G10L-015/22

International Patent Class (Additional): G06F-017/28; G10L-015/00;
G10L-015/18

File Segment: EPI; EngPI

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-100788

(43)Date of publication of application : 13.04.2001

(51)Int.Cl.

G10L 15/22

G06F 17/28

G10L 15/18

G10L 15/00

(21)Application number : 11-277742

(71)Applicant : SONY CORP

(22)Date of filing : 30.09.1999

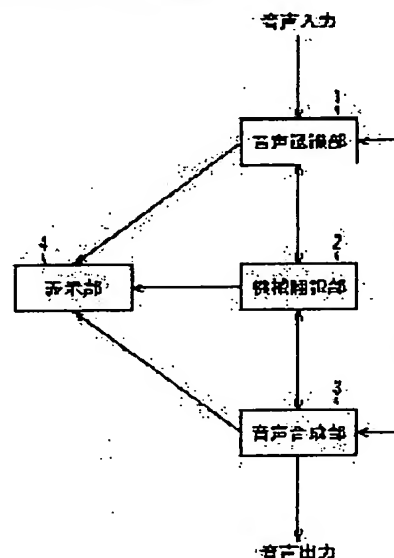
(72)Inventor : OGAWA HIROAKI
OMOTE MASANORI
HONDA HITOSHI
TSUTSUMI HIRONAGA

(54) SPEECH PROCESSOR, SPEECH PROCESSING METHOD AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To easily apply a high-grade natural language processing to speech.

SOLUTION: A speech recognition part 1 recognizes an input speech, and a machine translation part 2 performs machine translation of the speech recognition result. If recognizing it in the middle of the processing of the voice recognition result that a word included in the voice recognition result has been erroneously recognized, the machine translation part 2 requests homonyms of the word to the voice recognition part 1. In this case, the speech recognition part 1 recognizes the input speech again and outputs the speech recognition result including requested homonyms to the machine translation part 2.



音声翻訳システム

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Spec

図記号	図記号
(G1) In(C1)	P1
G10L 15/22	G10L 3/00
G06F 17/28	G06F 15/38
G10L 15/18	G10L 3/00
15/00	651C
	651E

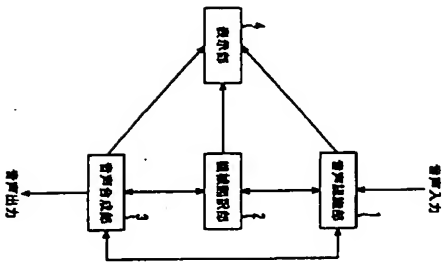
審査請求 未請求 請求項の数11 OL (全16頁)

(21) 出願番号	特願平11-277742	(71) 出願人	00002185 ソニー株式会社
(22) 出願日	平成11年9月30日 (1999.9.30)	(72) 発明者	東京都品川区北品川6丁目7番35号 小川 浩明 東京都品川区北品川6丁目7番35号 株式会社内 豊 稔則 (73) 発明者 東京都品川区北品川6丁目7番35号 ソニー株式会社内 (74) 代理人 100082131 弁護士 橋本 健雄

(71) 出願人
00002185
ソニー株式会社
(72) 発明者
東京都品川区北品川6丁目7番35号
小川 浩明
東京都品川区北品川6丁目7番35号
株式会社内
豊 稔則
(73) 発明者
東京都品川区北品川6丁目7番35号
ソニー株式会社内
(74) 代理人
100082131
弁護士 橋本 健雄

最終頁に続く

- (54) 【発明の名称】 音声処理装置および音声処理方法、並びに記録媒体
- (57) 【要約】
- 【課題】 音声に対して、容易に、高度な自然音処理を施す。
- 【解決手段】 音声認識部1において、入力音声の音声認識結果、機械翻訳部2において、その音声認識結果が機械翻訳される。機械翻訳部2は、音声認識結果の処理中に、その音声認識結果に含まれる単語が翻訳されたものであると認められる場合、その単語の同音異義語を、音声認識部1に要求する。この場合、音声認識部1は、入力音声を、再度音声認識し、要求された同音異義語を含めた音声認識結果を、機械翻訳部2に出力する。



音声認識システム

【特許請求の範囲】

- 【請求項1】 音声の音声認識する音声認識手段と、前記音声認識手段による音声認識結果を対象として自然音処理を行う自然音処理手段とを備える音声処理装置であって、
- 前記自然音処理手段は、前記自然音処理の途中結果に基づいて、前記音声認識手段に対して、所定の要求を行う要求手段を含み、
- 前記音声認識手段は、前記要求手段による要求に応じた処理を行うことを特徴とする音声処理装置。
- 【請求項2】 前記要求手段は、前記音声認識結果に含まれる単語の他の候補を要求し、
- 前記音声認識手段は、前記他の候補を、前記自然音処理手段に供給することを特徴とする請求項1に記載の音声処理装置。
- 【請求項3】 前記音声認識手段は、前記音声認識結果に含まれる単語の同音異義語を、前記自然音処理手段に供給することを特徴とする請求項2に記載の音声処理装置。
- 【請求項4】 前記要求手段は、前記音声認識結果に含まれる単語に、複数の候補が存在する場合に、確からしい候補の決定を要求することを特徴とする請求項1に記載の音声処理装置。
- 【請求項5】 前記音声認識手段は、前記複数の候補の中から、確からしい候補の決定を要求された場合、その複数の候補のフロンティアに基づいて、確からしい候補を決定することを特徴とする請求項4に記載の音声処理装置。
- 【請求項6】 前記要求手段は、音声認識の結果とする語彙の増加を要求することを特徴とする請求項1に記載の音声処理装置。
- 【請求項7】 前記要求手段は、前記音声認識結果に含まれる単語の他の候補の要求、前記音声認識結果に含まれる単語についての複数の候補の中からの確からしい候補の決定の要求、または音声認識の対象とする語彙の増加の要求のうち、少なくとも一つ以上を行うことを特徴とする請求項1に記載の音声処理装置。
- 【請求項8】 前記要求手段は、前記自然音処理の途中で、複数の処理結果の候補が得られた場合、確からしい候補を決定するための音響的な評価値を要求することとを特徴とする請求項1に記載の音声処理装置。
- 【請求項9】 前記自然音処理手段は、前記音声認識結果に対する返答の生成、または前記音声認識結果の機械翻訳を行うことを特徴とする請求項1に記載の音声処理装置。
- 【請求項10】 音声の音声認識する音声認識システムと、

- 前記自然音処理システムは、前記自然音処理の途中結果に基づいて、前記音声認識システムに対して、所定の要求を行う要求システムを含み、
- 前記音声認識システムにおいて、前記要求システムにおける要求に応じた処理を行うことを特徴とする音声処理方法。
- 【請求項11】 音声、コンピュータに処理させるプログラムが記録されている記録媒体であって、
- 前記音声の音声認識する音声認識システムと、
- 前記音声認識システムにおける音声認識結果を対象として自然音処理を行う自然音処理システムとを備えるプログラムが記録されており、
- 前記自然音処理システムは、前記自然音処理の途中結果に基づいて、前記音声認識システムに対して、所定の要求を行う要求システムを含み、
- 前記音声認識システムにおいて、前記要求システムにおける要求に応じた処理を行うことを特徴とする記録媒体。
- 【発明の詳細な説明】
- 【0001】
- 【発明の属する技術分野】 本発明は、音声処理装置および音声処理方法、並びに記録媒体に関し、特に、例えば、入力された音声に対して、容易に、高度な自然音処理を施すことができるようにする音声処理装置および音声処理方法、並びに記録媒体に関する。
- 【0002】
- 【従来の技術】 例えば、ある言語による音声、他の言語による音声に翻訳する音声認識システムでは、入力された音声を音声認識する音声認識装置を必須の構成要素として、その音声認識装置による音声認識結果が、自然音処理、即ち、機械翻訳される。
- 【0003】
- 【発明が解決しようとする課題】 ところで、上述のような音声認識システムでは、音声認識装置において、音声を機械翻訳すると、その翻訳した音声認識結果が機械翻訳されるため、その翻訳結果も翻訳したものとなる。
- 【0004】 ところで、音声認識装置において、高精度の音声認識を行うことにより、翻訳が行われる精度を低くすることができ、この場合、多くの語彙を対象とした、いわば負荷の重い音声認識処理を行う必要があり、処理に時間を要することになる。また、この場合、高速なハードウェアを用いることとすると、装置が高コスト化することになる。さらに、高精度の音声認識を行うには、一般に、幾つかの単語をまとめた単位で処理する必要があるが、このように、幾つかの単語をまとめた形で、音声認識結果が出力されると、その後の機械翻訳装置においては、その取り扱いが面倒になる。
- 【0005】 そこで、音声認識結果の尤度が低い場合等においては、発話者に、再度発話を行うように催促したり、音声認識結果が正しいかどうかを確認する方法があ

るが、そのような催促や強制を促すメッセージの生成等の処理管理が必要となり、システムが複雑化する。さらに、この場合、最終的な取扱い結果を得るまでに時間を要することとなり、ユーザ（発話者）の負担も増加することになる。

【0006】本発明は、このような状況に鑑みてなされたものであり、音声に対して、容易に、高度な自然言語処理を施すことができるようにするものである。

【0007】

【課題を解決するための手段】本発明の音声処理装置は、自然言語処理手段が、自然言語処理の途中結果に基づいて、音声認識手段に対して、所定の要求を行う要求手段を含み、音声認識手段が、要求手段による要求に応じた処理を行うことを特徴とする。

【0008】要求手段には、音声認識結果に含まれる語句の他の候補を要求させ、音声認識手段には、他の候補を、自然言語処理手段に供給させることができる。

【0009】音声認識手段には、音声認識結果に含まれる語句の同音異義語を、自然言語処理手段に供給させることができる。

【0010】要求手段には、音声認識結果に含まれる語句に、複数の候補が存在する場合に、適切な候補の決定を要求させることができる。

【0011】音声認識手段には、複数の候補の中から、適切な候補の決定を要求された場合、その複数の候補のプログラムに基づいて、適切な候補を決定させることができる。

【0012】要求手段には、音声認識の対象とする語彙の増加を要求させることができる。

【0013】また、要求手段には、音声認識結果に含まれる語句の他の候補の要求、音声認識結果に含まれる語句についての複数の候補の中から適切な候補の決定の要求、または音声認識の対象とする語彙の増加の要求のうちの、少なくとも1つ以上を行わせることができる。

【0014】さらに、要求手段には、自然言語処理の途中で、複数の処理結果の候補が得られた場合、適切な候補を決定するための音響的な評価値を要求させることができる。

【0015】自然言語処理手段には、音声認識結果に対する返答の生成、または音声認識結果の候補取扱いを行わせることができる。

【0016】本発明の音声処理方法は、自然言語処理ステップが、自然言語処理の途中結果に基づいて、音声認識ステップに対して、所定の要求を行う要求ステップを含み、音声認識ステップにおいて、要求ステップにおける要求に応じた処理を行うことを特徴とする。

【0017】本発明の処理装置は、自然言語処理ステップが、自然言語処理の途中結果に基づいて、音声認識ステップに対して、所定の要求を行う要求ステップを含

60

み、音声認識ステップにおいて、要求ステップにおける要求に応じた処理を行うことを特徴とする。

【0018】本発明の音声処理装置および音声処理方法、並びに記録媒体においては、自然言語処理の途中結果に基づいて、所定の要求が行われ、その要求に応じた処理が行われる。

【0019】

【発明の実施の形態】図1は、本発明を適用した音声認識システム（システム）とは、複数の装置が物理的に集合したものをいい、各装置の装置が同一筐体中にあるか否かは問わない）の一実施の形態の構成例を示している。【0020】この音声認識システムでは、音声が入力されると、その音声の対応が出力されるようになっている。

【0021】即ち、音声認識部1には、例えば日本語などによる音声が入力されるようになっており、音声認識部1は、入力された音声や音声認識し、その音声認識結果としてのテキスト、その他付随する情報を、機械翻訳部2や、表示部4などに出力する。

【0022】機械翻訳部2は、音声認識部1が出力する音声認識結果を解析し、入力された音声の音素以外の、例えば英語に機械翻訳し、その翻訳結果としてのテキスト、その他付随する情報を、音声合成部3や、表示部4などに出力する。音声合成部3は、機械翻訳部2など出力に基づいて音声合成処理を行い、入力された音声の、他の音素への翻訳結果としての合成音を出力する。

【0023】表示部4は、例えば、液晶ディスプレイ等で構成され、音声認識部1による音声認識結果や、機械翻訳部2による機械翻訳結果等を表示する。

【0024】以上のように構成される音声認識システムにおいては、入力された音声や、音声認識部1で音声認識され、機械翻訳部2に供給される、機械翻訳部2では、音声認識部1による音声認識結果が機械翻訳され、音声合成部3に供給される。音声合成部3では、機械翻訳部2からの翻訳結果に対応する合成音が生成されて出力される。

【0025】次に、図2は、図1の音声認識部1の構成例を示している。

【0026】ユーザの発話、マイク11に入力され、マイク11では、その発話、電気信号としての音声信号に変換される。この音声信号は、AD（Analog/Digital）変換部12に供給される。AD変換部12では、マイク11からのアナログ信号である音声信号がサンプリング、量子化され、デジタル信号である音声データに変換される。この音声データは、特徴抽出部13に供給される。

【0027】特徴抽出部13は、AD変換部12からの音声データについて、適切なフレームごとに、例えば、スペクトルや、線形予測誤差、ケプストラム係数、線スペクトル対等の特徴パラメータを抽出し、特徴量パッ

60

ア14およびマッピング部15に供給する。特徴量パッファ14では、特徴抽出部13からの特徴パラメータが一時記憶される。

【0028】マッピング部15は、特徴抽出部13からの特徴パラメータ、または特徴量パッファ14に記憶された特徴パラメータに基づき、音響モデルデータベース16、辞書データベース17、および文法データベース18を必要に応じて参照しながら、マイク11に入力された音声（入力音声）を認識する。

【0029】即ち、音響モデルデータベース16は、音声認識する音声の音素における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここで、音響モデルとしては、例えば、HMM（Hidden Markov Model）などを用いることができる。辞書データベース17は、認識対象の各単語について、その発音に関する情報が記述された単語辞書を記憶している。文法データベース18は、辞書データベース17の単語辞書に登録されている各単語が、どのように連綴する（つながる）かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文法自由文法（CFG）や、統計的な単語連綴率（N-gram）などに基づく規則を用いることができる。

【0030】マッピング部15は、辞書データベース17の単語辞書を参照することにより、音響モデルデータベース16に記憶されている音響モデルを接続することによって、マイク11に入力された音声や音素を、文法に、マッピング部15は、幾つかの単語モデルを、文法データベース18に記憶された文法規則を参照することによって接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、例えば、HMM法等によって、マイク11に入力された音声や音素を認識する。

【0031】そして、マッピング部15による音声認識結果は、例えば、テキスト等で出力される。

【0032】なお、マッピング部15は、機械翻訳部2から、後述する要求信号を受信することができるようになっており、要求信号を受信した場合、その要求信号による要求に応じた処理を行うようにもなっている。この場合、マッピング部15は、一度処理した特徴パラメータが再び必要となるときは、特徴量パッファ14に記憶された特徴パラメータを用いて処理を行うようになっている。これにより、ユーザに、再度発話をしなくても済むようになる。

【0033】次に、図3は、図1の機械翻訳部2の構成例を示している。

【0034】テキスト解析部21には、音声認識部1が出力する音声認識結果としてのテキストが、機械翻訳部21は、辞書データベース24や解析用文法データベース25を参照しながら、そのテキストを解析する。【0035】即ち、辞書データベース24には、各単語

60

の表記や、解析用文法を適用するために必要な品詞情報などが記述された単語辞書が記憶されている。また、解析用文法データベース26には、単語辞書に記述された各単語の情報に基づいて、単語辞書に関する制約等が記述された解析用文法規則が記憶されている。そして、テキスト解析部21は、その単語辞書や解析用文法規則に基づいて、そこに入力されるテキスト（入力テキスト）の形態素解析や、構文解析等を行い、その入力テキストを構成する単語や構文の情報等の言語情報を抽出する。

ここで、テキスト解析部21における解析方法としては、例えば、正規文法や、文法自由文法、統計的な単語連綴率を用いたものなどがある。

【0036】テキスト解析部21で得られた入力テキストの解析結果としての言語情報は、言語交換部22に供給される。言語交換部22は、言語交換データベース26を参照し、入力テキストの言語の言語情報を、翻訳結果の言語の言語情報に変換する。

【0037】即ち、言語交換データベース26には、入力言語（言語交換部22への入力言語）の言語情報から、出力言語（言語交換部22からの出力言語）の言語情報への変換パターンや、入力言語と出力言語との対応関係およびその対応関係と入力言語との間の類似度の計算に用いられるシソーラス等の、言語情報を変換するための言語交換データが記憶されている。そして、言語交換部22では、このような言語交換データに基づいて、入力テキストの言語の言語情報が、出力言語の言語情報に変換される。

【0038】なお、言語交換部22は、その処理の途中結果に基づいて、その処理を精確良く行うために必要な情報があるとき、その情報を要求する要求信号を、音声認識部1（のマッピング部15）に出力し、その情報を得る、精確の良い処理を行うようになっている。

【0039】言語交換部22で得られた出力言語の言語情報は、変換結果修正部29に供給される。変換結果修正部29は、言語交換部22からの出力言語を評価し、その評価結果が所定の基準を満たしている場合には、言語交換部22からの言語情報を、そのままテキスト生成部23に供給する。

【0040】一方、言語交換部22からの言語情報の評価結果が所定の基準を満たしていない場合には、変換結果修正部29は、その言語情報を、変換結果修正部29に一時記憶させるとともに、音声認識結果の他の候補を要求する要求信号を、音声認識部1（のマッピング部15）に出力する。そして、音声認識部1が、変換結果修正部29からの要求信号に応じて、音声認識結果の他の候補を出力し、それが、テキスト解析部21および言語交換部22で処理され、出力言語の言語情報とされ、変換結果修正部29に供給されると、変換結果修正部29は、その新たに受信した言語情報と、変換結果修正部29に記憶された言語情報とを比較し、入力テキスト

60

の処理結果として、より適切な方を選択して、テキスト生成部23に出力する。

【0041】テキスト生成部33は、辞書データベース27および生成用文法データベース28を参照することにより、変換結果修正部38からの出力言語の言語情報から、入力テキストを出力言語に翻訳したテキストを生成する。

【0042】即ち、辞書データベース27には、出力言語の文を生成するために必要な単語の品詞や活用形等の情報で記述された単語辞書が記憶されており、また、生成用文法データベース28には、出力言語の文を生成するために必要な単語の活用規則や語順の制約等の生成用文法規則が記憶されている。そして、テキスト生成部23は、これらの単語辞書および生成用文法規則に基づいて、言語変換部22からの言語情報を、テキストに変換して出力する。

【0043】次に、図4は、図1の音声合成部3の構成例を示している。

【0044】テキスト解析部31には、機械翻訳部2が出力する翻訳結果としてのテキストが、音声合成処理の対象として入力されるようになっている。テキスト解析部31は、辞書データベース34や解析用文法データベース35を参照しながら、そのテキストを解析する。

【0045】即ち、辞書データベース34には、各単語の品詞情報や、読み、アクセント等の情報が記憶された単語辞書が記憶されており、また、解析用文法データベース35には、辞書データベース34の単語辞書に記述された単語について、単語連綴に関する制約等の解析用文法規則が記憶されている。そして、テキスト解析部31は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部32で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の韻律情報や、各単語の発音等の音韻情報などがあ

る。

【0046】テキスト解析部31で得られた情報は、規則合成部32に供給され、規則合成部32では、音素片データベース36を用いて、テキスト解析部31に入力されたテキストに対応する合成音の音素データ（ディジタルデータ）が生成される。

【0047】即ち、音素片データベース36には、例えば、CV (Consonant, Vowel) や、VCV、CVC等の形で音素片データが記憶されており、規則合成部32は、テキスト解析部31からの情報に基づいて、必要な音素片データを選択し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト解析部31に入力されたテキストに対応する合成音の音素データを生成する。

【0048】この音素データは、DA変換部33に供給され、ここで、アナログ信号としての音声信号に変換される。この音声信号は、図示せぬスピーカに供給され、これにより、テキスト解析部31に入力されたテキストに対応する合成音が出力される。

【0049】次に、図5は、図3の言語変換部22および言語変換データベース26の構成例を示している。

【0050】言語変換部22は、ワッペン部41から構成されており、ワッペン部41は、言語変換データベース26を参照することで、テキスト解析部31から供給される入力テキストの言語情報を、出力言語の言語情報に変換する。

【0051】即ち、図5の實施の形態では、言語変換データベース26は、シソーラス辞書42およびワッペン部43から構成されている。シソーラス辞書42は、例えば、図6に示すように、単語およびその概念（意味）が階層的に分類されたシソーラスを記憶している。なお、図6においては、長方形で囲んである記載が概念を表し、楕円形で囲んである記載が単語を表して

いる。また、概念を表す長方形に記載されている番号は、シソーラスにおいて、その概念を特定するための分類番号である。

【0052】図6に戻り、ワッペン部43には、日本語の文のパターンと、その英語訳とを対応付けたワッペンが登録されている。なお、ワッペン部43では、日本語文のパターンは、必要に応じて変換（図6においては、XやY）を用いて記述されている。また、変換には番号が付けられている。この番号は、図6に示したシソーラス上の分類番号を表す。

【0053】以上のように構成されるワッペン部41では、ワッペン部43に登録されているワッペンから、入力テキストと最も類似する（距離の近い）日本語文のパターンを有するものが選択される。即ち、ワッペン部41は、ワッペン部43の各ワッペンと日本語文のパターンと、入力テキストとの距離を求め、その距離を最も短くするワッペンとを選択する。さらに、その選択されたワッペン部41（以下、適宜、選択ワッペン部という）の日本語文のパターンの中の変換に対応する語句（以下、適宜、対応語句という）が、入力テキストから抽出され、その対応語句と、選択ワッペン部の変換が属する概念との間の意味的な距離（以下、適宜、単語間距離という）が求められる。

【0054】ここで、選択ワッペン部の変換と、その対応語句との単語間距離とは、例えば、シソーラスにおいて、その変換のノードから、対応語句のノードへ移動するための最小移動数（変換のノードから、対応語句のノードへの最短のパスを構成するワッペン部の数）などを用いることができる。なお、選択ワッペン部により、複数の変換が存在する場合には、その複数の変換それぞれ

について、単語間距離が計算される。

【0055】ワッペン部41は、選択ワッペン部を決定し、さらに、その選択ワッペン部における変換についての単語間距離を求めると、選択ワッペン部、対応語句、および単語間距離とともに、選択ワッペン部における日本語文のパターンと入力テキストとの距離（これは、上述のように、選択ワッペン部を決定するときに求められる）（以下、適宜、パターン間距離という）を、変換結果修正部29に出力する。

【0056】具体的に、例えば、入力テキストが「金曜を使った」であった場合において、日本語文「X (1. 5621) を使う」を有するワッペン部が、選択ワッペン部として決定（選択）されたとして、その選択ワッペン部、対応語句「金曜」、入力テキスト「金曜を使った」と日本語文「X (1. 5621) を使う」とのパターン間距離、並びに変換X (1. 5621) とその対応語句「金曜」との単語間距離が、変換結果修正部29に出力される。

【0057】なお、ワッペン部41は、基本的には、上述のように、入力テキストとのパターン間距離を最も短くするワッペン部を、そのまま、選択ワッペン部として決定するが、後述するように、2つなどの複数のワッペン部が、選択ワッペン部の候補として得られ、最終的な選択ワッペン部を決定することが困難な場合には、音声認識部1に対して、より厳かしいワッペン部を、選択ワッペン部として決定するに必要となる情報（要求する要求信号を出力し、その要求した情報）が、音声認識部1から供給されるのを待つ。選択ワッペン部を決定するようになっている。

【0058】次に、図7のフローチャートを参照して、図3の機械翻訳部2の動作について、さらに説明する。

【0059】機械翻訳部2では、音声認識部1から音声認識結果としての入力テキストを受信すると、ステップ3.1において、変換結果パツツフ3.0の記憶内容がリフされ、ステップ3.2に進み、テキスト解析部2.1は、入力テキストの解析を行う。テキスト解析部2.1による解析結果は、言語変換部2.2に供給され、言語変換部2.2は、ステップ3.3において、図6で説明したような選択ワッペン部の決定等を行い、その選択ワッペン部を用いて、入力テキストの言語情報を出力テキストの言語情報に変換し、その変換結果としての、上述した選択ワッペン部、パターン間距離、対応語句、および単語間距離を、変換結果修正部2.9に出力する。

【0060】変換結果修正部2.9は、ステップ3.4において、言語変換部2.2からの出力テキストの言語情報（選択ワッペン部、パターン間距離、対応語句および単語間距離）を、変換結果パツツフ3.0に供給して記憶させ、ステップ3.5に進み、言語変換部2.2から供給された単語間距離が、所定の基準値以下であるかどうかを判定する。ステップ3.6において、単語間距離が、所定

の基準値以下であると判定された場合、即ち、選択ワッペン部の変換が属する概念と、その変換についての入力テキストの対応語句との意味的な距離が近く、従って、入力テキストの対応語句が、正しい音声認識結果であると推定される場合、ステップ3.6に進み、変換結果修正部2.9は、ステップ3.4で変換結果パツツフ3.0に記憶させた出力言語の言語情報を、テキスト生成部23に出力する。そして、テキスト生成部23では、入力テキストを出力言語に翻訳したテキストが生成され、処理を終了する。

【0061】一方、ステップ3.6において、単語間距離が、所定の基準値以下でない判定された場合、即ち、選択ワッペン部の変換が属する概念と、その変換についての入力テキストの対応語句との意味的な距離が遠く、従って、入力テキストの対応語句が、誤った音声認識結果であると推定される場合（ここでは、音響的には正しいが、意味的に誤っている）と推定される場合）、ステップ3.7に進み、変換結果修正部2.9は、対応語句について、例えば、その対応語句と同音異義語の他の候補を要求する要求信号を、音声認識部1に出力する。

【0062】この場合、音声認識部1では、特徴量パツツフ1.4に記述された特徴パラメータを用いて音声認識が実行され、これにより、対応語句と同音異義語の他の候補が、機械翻訳部2に供給される。なお、対応語句と同音異義語を求めた出力テキスト認識は、例えば、音声認識部1の辞書データベース1.7に、各種の同音異義語を記憶させておくことで行うことが可能である。ここで、本明細書においては、同音異義語には、厳密な意味での同義異義語だけでなく、音響的に近い語も含められるものとする。

【0063】対応語句と同音異義語の他の候補が、音声認識部1から機械翻訳部2に供給されると、ステップ3.8において、テキスト解析部2.1および言語変換部2.2は、入力テキストの対応語句を、その同音異義語の他の候補に置き換えたもの（以下、適宜、置換テキストという）を対象に処理を行う。そして、その処理結果は、変換結果修正部2.9に出力される。

【0064】なお、対応語句と同音異義語の他の候補が、複数存在する場合には、音声認識部1から機械翻訳部2に対しては、その複数の候補が供給される。この場合、機械翻訳部2では、その複数の候補それぞれについて、置換テキストが作成されて処理される。

【0065】以上のようにして、変換結果修正部2.9は、言語変換部2.2から、置換テキストを交換した出力言語の言語情報を受信すると、ステップ3.9において、その言語情報と、変換結果パツツフ3.0に記憶されている言語情報のうち、より適切なものを選択する。即ち、変換結果修正部2.9は、置換テキストを交換した出力言語の言語情報と、変換結果パツツフ3.0に記憶されている言語情報のうち、その言語情報に含まれる単語間距離

ンゾ部16は、通常は、音響モデルデータベース16、辞書データベース17、および文法データベース18だけを用いて処理を行うが、機械翻訳部2からの要求信号に応じて、分野別辞書群19の必要な分野別辞書を参照して処理を行うようになっている。

[0094] 次に、図12は、音声認識部1が図11に示したように構成される場合の、図1の機械翻訳部2の構成例を示している。なお、図中、図3における場合と対応する部分については、同一の符号を付しており、以下では、その説明は、適宜省略する。即ち、図12の機械翻訳部2は、分野指定部51が新たに設けられている。図3における場合と同様に構成されている。

[0095] 分野指定部51には、テキスト解析部21から、入力テキストの解析結果が供給されるとともに、変換結果修正部29が出力する要求信号が供給されるようになっている。そして、分野指定部51は、テキスト解析部21からの解析結果に基づいて、即ち、それまでに入力された音声の音声認識結果の解析結果に基づいて、ユーザの発話内容の分野や話題を指定し、変換結果修正部29から要求信号を受信した場合には、指定した分野や話題に対応する分野別辞書を追加して処理を行うように要求する要求信号を、音声認識部1に供給するようになっている。

[0096] ここで、発話から、その内容の分野や話題を指定する方法については、例えば、「自然言語理解システムにおける分野推論方式」、別所文人、岩間成人、戸部典孝、福村好美、電子情報通信学会論文誌E-11 J01-011 No.6 pp.1317-1327などに、その詳細が記載されている。

[0097] 次に、図13のフローチャート参照して、図12の機械翻訳部2の動作について説明する。

[0098] 図12の機械翻訳部2では、ステップS41乃至S44において、図7のステップS1乃至S4における場合とそれぞれ同様の処理が行われる。

[0099] そして、ステップS44の処理後は、ステップS46に進み、変換結果修正部29は、音節変換部22から供給されたパターン間距離が、所定の基準値以下であるかどうかを判定する。ステップS46において、パターン間距離が、所定の基準値以下であると判定された場合、即ち、音声認識結果と、選択パターン間距離に近づくように、音声認識部1(図11)において、分野別辞書群19を用いながらも、正しい音声認識結果が得られていると推測される場合、ステップS46に進み、変換結果修正部29は、ステップS44で変換結果修正部29に記憶させた出力言語の音節情報と、テキスト生成部23に出力する。そして、テキスト生成部23では、入力テキストを出力言語に翻訳したテキストが生成され、処理を終了する。

[0100] 一方、ステップS46において、パターン

間距離が、所定の基準値以下でないと判定された場合、即ち、音声認識結果と、選択パターン間距離に記憶された日本語文との距離が遠く、従って、音声認識部1(図11)において、分野別辞書群19をさらに用いて処理を行わないと、正しい音声認識結果が得られないと推測される場合、ステップS47に進み、変換結果修正部29は、分野別辞書群19を用いた音声認識を要求する要求信号を、分野指定部51に供給する。

[0101] 分野指定部51は、テキスト解析部21の出力を参照することで、発話内容の話題や分野を指定しており、変換結果修正部29から要求信号を受信すると、指定した分野や話題に対応する分野別辞書を追加して処理を行うように要求する要求信号を、音声認識部1に供給する。

[0102] 即ち、分野指定部51は、例えば、発話内容の話題が旅行に関することであると指定した場合に、観光地の地名が登録されている分野別辞書を追加し、これにより、音声認識の対象語彙として、観光地の地名を含めて処理を行うように要求する要求信号を、音声認識部1に供給する。

[0103] この場合、音声認識部1では、特徴量ベクトル14に記憶された特徴パラメータを用い、要求信号に記した分野や話題に関する単語(即ち)の情報が登録されている分野別辞書をさらに参照して音声認識が行われ、これにより、音声認識の対象語彙を、いわば拡張した音声認識が行われる。そして、その音声認識結果は、音声認識部1から機械翻訳部2に供給される。

[0104] 上述のようにして、新たな音声認識結果が、音声認識部1から供給されると、機械翻訳部2では、ステップS48において、テキスト解析部21および音節変換部22は、その新たな音声認識結果としての入力テキストを対応に処理を行う。そして、その処理結果は、変換結果修正部29に出力される。

[0105] 以上のようにして、変換結果修正部29は、言語変換部22から、新たな音声認識結果を交換した出力言語の音節情報を受信すると、ステップS49において、その音節情報と、変換結果修正部29に記憶されている音節情報のうち、より適切なものを選択する。即ち、変換結果修正部29は、新たな音声認識結果を交換した出力言語の音節情報と、変換結果修正部29に記憶されている音節情報のうち、パターン間距離が小さいものを選択する。

[0106] そして、ステップS46に進み、変換結果修正部29は、選択した音節情報と、テキスト生成部23に出力し、以下、上述した場合と同様の処理が行われ、処理を終了する。

[0107] 次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、

専用のハードウェアとしての音声翻訳システムに組み込まれているコンピュータや、汎用のコンピュータ等にインストールされる。

[0108] そして、図14を参照して、上述した一連の処理を実行するプログラムをコンピュータにインストールし、コンピュータによって実行可能な状態とするために用いられる、そのプログラムが記録されている記録媒体について説明する。

[0109] プログラムは、図14(A)に示すように、コンピュータ101に内蔵されている記録媒体としてのハードディスク102や半導体メモリ103に予め記録しておくことができる。

[0110] あるいはまた、プログラムは、図14(B)に示すように、コンパクトディスク111、CD-R(DCompact Disc Read Only Memory)112、MO(Magneto optical)ディスク113、DVD(Digital Versatile Disc)114、磁気ディスク115、半導体メモリ116、デジタル放送放送用の人工衛星122を介して、コンピュータ101に無線で送ったり、LAN(Local Area Network)、インターネットといったネットワーク131を介して、コンピュータ123に有線で送渡し、コンピュータ101において、内蔵するハードディスク102などにインストールすることができ、

[0111] なお、プログラムは、上述したような記録媒体からコンピュータにインストールする他、図14(C)に示すように、ダウンロード121から、デジタル放送放送用の人工衛星122を介して、コンピュータ101に無線で送ったり、LAN(Local Area Network)、インターネットといったネットワーク131を介して、コンピュータ123に有線で送渡し、コンピュータ101において、内蔵するハードディスク102などにインストールすることができ、

[0112] ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは別個に実行される処理(例えば、並列処理あるいはサブタスクによる処理)も含むものである。

[0113] また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

[0114] 次に、図15は、図14のコンピュータ101の構成例を示している。

[0115] コンピュータ101は、図16に示すように、CPU(Central Processing Unit)142を内蔵している。CPU142には、バス141を介して、入力インタフェース145が接続されており、CPU142は、入力ポートやバス等で構成される入力部147が操作されることにより指令入力されると、それにしたがっ

て、図14(A)の半導体メモリ103に対応するROM(Read Only Memory)143に格納されているプログラムを実行する。あるいは、また、CPU142は、ハードディスク102に格納されているプログラム、衛星122若しくはネットワーク131から転送され、通信部148で受信されてハードディスク102にインストールされたプログラム、またはダイナミック149に格納されたフラッシュメモリ111、CD-ROM112、ハードディスク113、DVD114、若しくは磁気ディスク115から読み出されてハードディスク102にインストールされたプログラムを、RAM(Random Access Memory)144にロードして実行する。そして、CPU142は、その処理結果を、例えば、入出力インタフェース145を介して、LCD(Liquid Crystal Display)等で構成される表示部146に、必要に応じて出力する。

[0116] 以上のように、機械翻訳部2において、その処理の途中結果に基づいて、音声認識部1に対して、所定の要求を行い、音声認識部1において、その要求に応じた処理を行うようにしたので、音声に対して、容易に、高度な自然言語処理を施すことが可能となる。

[0117] 即ち、この場合、音声認識部1では、比較的簡易な音声認識処理を行い、機械翻訳部2において、その音声認識結果の処理の最中に、疑問が生じたときや、新たな情報が必要になったときに、音声認識部1にその疑問を解決するための処理や必要な情報を要求することができ、その結果、機械翻訳部2では、音声に対して、容易に、高度な自然言語処理を施すことが可能となる。

[0118] また、この場合、発話者、再度発話を行うように促したり、音声認識結果が正しいかどうかを確認する必要はない。

[0119] なお、本実施の形態では、機械翻訳部2において、日本語のバグーンを有するパターンを用いて翻訳を行うようにしたが、翻訳は、その他、例えば、用例を用いて行うようにすることも可能である。

[0120] また、本実施の形態では、音声認識結果を機械翻訳部2に示したが、本発明は、音声認識結果に対する返答を生成して対話を行うための対話管理処理、その他の自然言語処理を行う装置に適用可能である。

[0121]

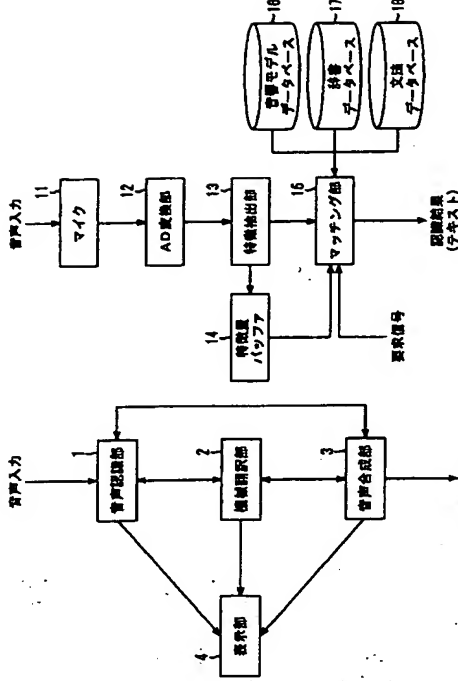
[発明の効果] 本発明の音声処理装置および音声処理方法、並びに記録媒体によれば、音声認識結果に対する自然言語処理の途中結果に基づいて、所定の要求が行われ、その要求に応じた処理が行われる。従って、音声に対して、容易に、高度な自然言語処理を施すことが可能となる。

[図面の簡単な説明]

[図1] 本発明を適用した音声翻訳システムの一実施の形態の構成例を示すブロック図である。

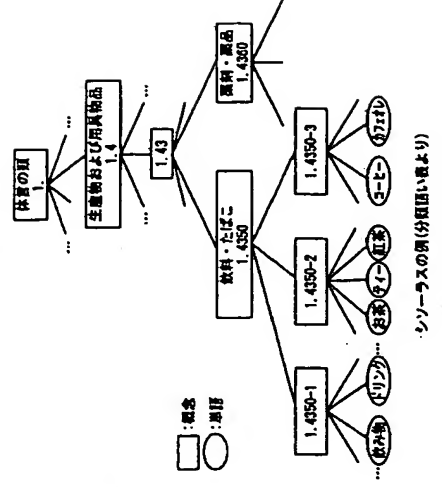
[図2] 音声認識部1の情報例を示すブロック図であ

【図1】



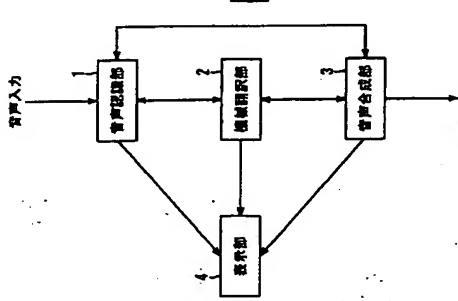
音声認識システム

【図6】



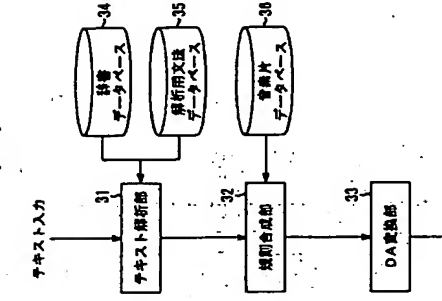
音声認識部1

【図2】



音声認識システム

【図4】

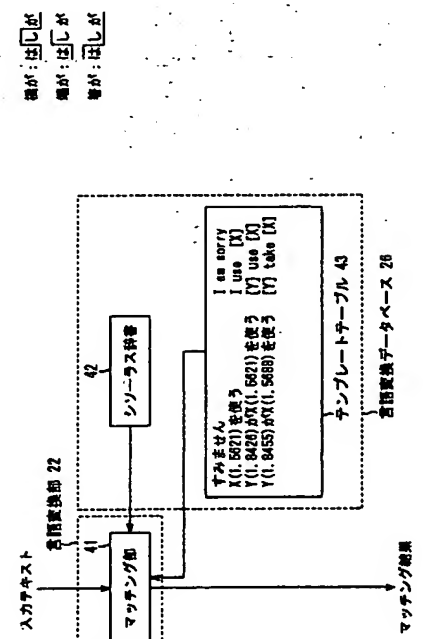


音声認識部1

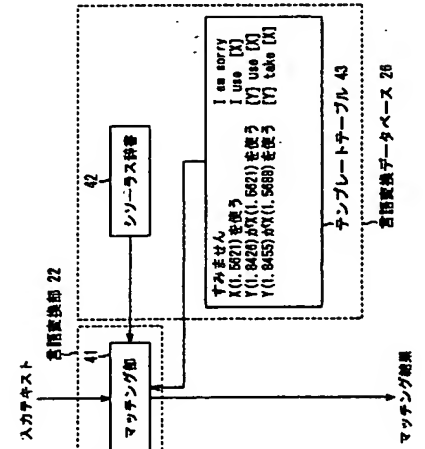
【符号の説明】

- 1 音声認識部, 2 機械翻訳部, 3 音声合成部, 4 表示部, 5 対話管理部, 11 マイク (マイクプロフォン), 12 AD変換部, 13 特徴抽出部, 14 特徴量バンプ, 15 マッチング部, 16 音声モデルデータベース, 17 辞書データベース, 18 文法データベース, 19 分詞別辞書, 21 テキスト解析部, 22 音素変換部, 23 テキスト生成部, 24 辞書データベース, 25 解析用文法データベース, 26 音素変換データベース, 27 辞書データベース, 28 生成用文法データベース, 29 変換結果修正部, 30 変換結果バンプ, 31 テキスト解析部, 32 規則生成部, 33 DA変換部, 34 辞書データベース, 35 解析用文法データベース, 42 音声認識部, 43 テンプレートデータベース, 51 分詞別辞書, 101 コンビュタ, 102 ハードディスク, 103 半導体メモリ, 111 フロッピーディスク, 112 CD-ROM, 113 MOディスク, 114 DVD, 115 磁気ディスク, 116 半導体メモリ, 121 ダウンロードサイト, 122 新島, 131 ネットワーク, 141 バス, 142 CPU, 143 ROM, 144 RAM, 145 入力インターフェース, 146 表示部, 147 入力部, 148 通信部, 149 ドライブ

【図9】

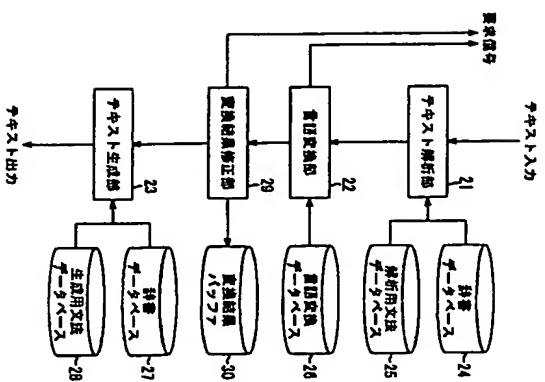


【図6】

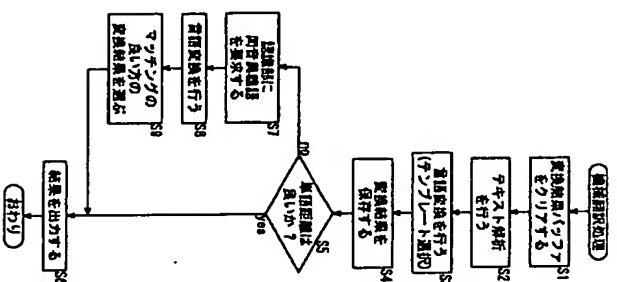


音声認識部1

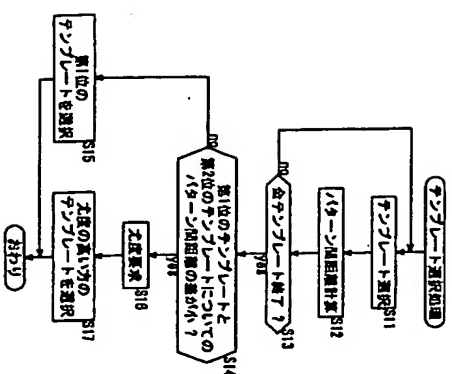
【圖 3】



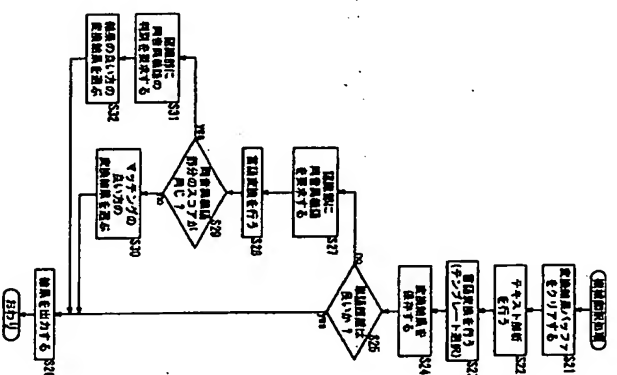
【圖 7】



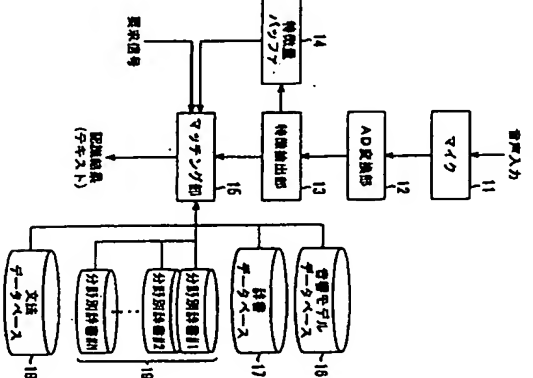
[88]



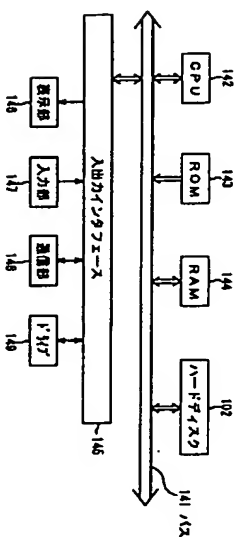
[0 1 0]



【圖 11】



[9 1 5]



【図12】

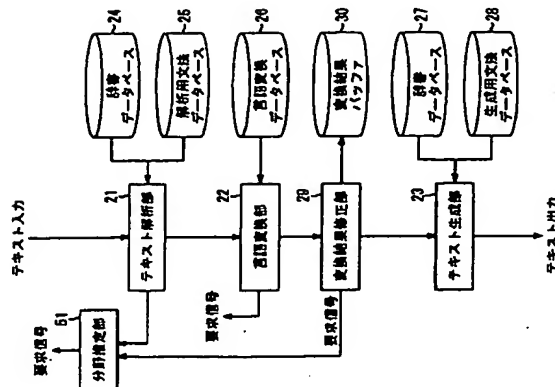
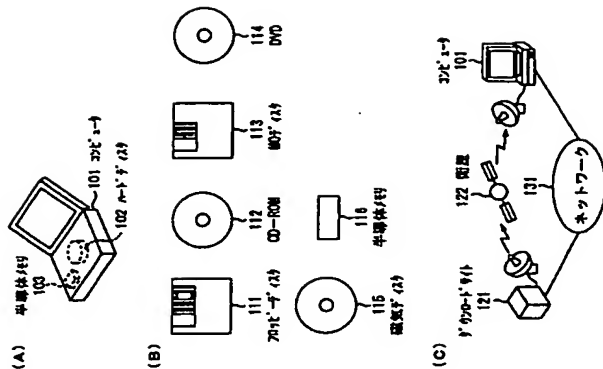


図12図部 2

【図14】



フロントページの続き

(72)発明者 本田 等
Fターム(参考) 5B091 AA15 BA03 CB12 CC32 CC02
E2A04
5D016 AA05 BB02 HH00 HH04 KK04
LL04 LL06 LL12
(72)発明者 包 洪良
9A001 BB06 DD16 HH14 HH17 HH18
KK46
東京都品川区北品川6丁目7番35号 ソニ
株式会社内
東京都品川区北品川6丁目7番35号 ソニ
株式会社内